

Veda Kailasam

✉ vedak2@illinois.edu | ☎ +1(682) 999-6778 | 📠 vedakailasam | 🌐 www.vedakailasam.com

EDUCATION

University of Illinois, Urbana Champaign

Master of Computer Science

Relevant Coursework: Using LLMs AKA ChatGPT, Computer Vision, Applied Machine Learning, Text Information Systems, Data Science Algorithms

GPA: 4.0/4.0

Aug 2023 - Dec 2024

Jawaharlal Nehru Technological University, Hyderabad, India

Bachelor of Technology in Computer Science & Engineering

GPA: 3.9/4.0

Aug 2017 - June 2021

SKILLS

- **Languages:** Python, SQL, Shell Scripting, HTML, JavaScript, PHP
- **Technologies:** GCP, AWS, Git, Jira, Docker, Kafka, Flask, Flutter, MongoDB, Linux
- **Artificial Intelligence:** LLMs, Gen AI, LLMOps, LangChain, LLamaIndex, Prompt Engineering, Accelerate, Knowledge Distillation, Quantization, GANs, Transformers, MLOps, NLP, Robotics, ML, Deep Learning, Vision, RecSys, Data Science,
- **Frameworks/Libraries:** Pyspark, PEFT, Elastic Search, Pandas, NumPy, Scikit-Learn, Keras, Tensorflow, PyTorch, Matplotlib, OpenCV, NLTK, Spacy, Neo4J, Streamlit

EXPERIENCE

Graduate Research and Teaching Assistant, Department of Computer Science

Champaign, IL

University of Illinois Urbana-Champaign

Jan 2024 - Present

- Collaborated with Professor David Dalpiaz on **CS 498: End-to-End Data Science**, focusing on curriculum development, student support, and the design of an autograder for programming assignments.
- Spearheaded research on personalized educational tools using **LLMs** and **generative AI**, significantly advancing data science education.
- Tailored educational content through re-ranking techniques in information retrieval, enhancing content personalization.

Machine Learning Engineer

Bengaluru, India

VerSe Innovation (Dailyhunt & Josh)

July 2021 - July 2023

- Enhanced Dailyhunt's recommendation engine through a sophisticated **two-tower model** with **Pyspark** and **NLP**, which significantly refined user profiles and content metadata, resulting in a **25%** improvement in top-k recommendations across English and Indic languages.
- Implemented advanced end-to-end content processing pipelines on Dailyhunt using **Vertex AI** and **GCP**, including **LDA**, **K-Means**, and **Locality-Sensitive Hashing** achieving a **30%** increase in user interaction and a **20%** improvement in content relevancy.
- Automated the grouping of news articles into similar topics using **LDA** and **K-Means**, captured engaging stories using advanced techniques such as **Locality-Sensitive Hashing (LSH)** and **knowledge graphs**, and fostered user engagement through **community detection** using **named entities**.
- Pioneered the development of an efficient and cost-effective **Named Entity Recognition (NER)** pipeline at Dailyhunt by integrating a **Conditional Random Field (CRF)** based model with a **BERT-based transformer**. This advanced approach achieved over **90%** accuracy in multilingual entity detection enhancing the quality of content categorization and search functionalities.
- Enhanced Dailyhunt's efficiency by **40%** by automating data pipelines and **PySpark jobs** on **DataProc clusters**. Integrated Dailyhunt's entire end-to-end content processing into a unified pipeline, "**Starfish**," which provides components like category, story ID, and topic ID for news articles within seconds
- Optimized Dailyhunt's performance by deploying scalable machine learning models on **GCP**, leveraging fine-tuned **large language models** to enhance **content personalization** and system responsiveness, resulting in a **25%** reduction in response time.

Artificial Intelligence Intern

Bengaluru, India

Wipro LTD.

Sep 2020 - May 2021

- Specialized in face recognition and **text detection** by utilizing Transformer-based models in **PyTorch** and **Keras**, developing robust face attribute classifiers and enhancing **OCR** capabilities, significantly improving image-based text processing and facial analysis accuracy by **40%**.
- Contributed to the engineering life-cycle, including designing high-quality AI infrastructure and data pipelines, writing **production code**, resulting in a **25%** improvement in system reliability.

PROJECTS

- **[Research Paper Project] Refine, Compress, Re-rank: Improving Listwise Re-ranking in Large Language Models** : Engineered a sophisticated multi-stage re-ranking pipeline leveraging **MonoBERT** to eliminate initial ordering biases, **FLAN-T5** for data compression, and **GPT models** for enhanced **re-ranking**, boosting **query relevance** by **15%** and reducing **operational costs** by **30%**, while eliminating sliding windows and initial ordering biases.
- **LegalLLM: Legal Document Summarization**: Developed a tool that leverages **fine-tuned LLMs** and **Retrieval-Augmented Generation (RAG)** to analyze and summarize legal documents, streamlining key term extraction and clause analysis for law firms and legal departments resulting in a **50%** reduction in time spent on manual document review.
- **StreamCLIP: Intelligent Video Search and Genre-Based Extraction** : Built "StreamClip," a video search and extraction tool powered by advanced AI models like **DistilBert**, **ResNet18**, **ViT**, and **CLIP**. It enables precise searches and genre-based extraction, revolutionizing video content discovery with a **50%** increase in search accuracy and **30%** reduction in extraction time
- **Enhancing Photo Organization through Face Recognition and Clustering**: Led a project that enhanced photo gallery organization by implementing face recognition and clustering technologies, achieving **97.8%** accuracy on a controlled dataset and **88.3%** on the expanded dataset, demonstrating robust model generalization.